AMENDMENTS TO THE CLAIMS:

Claims 1-28 (Cancelled)

29. (Currently amended) A method for a data processing system to efficiently cluster ~~of clustering~~ data points from a ~~dataset~~ dataset, the method comprising the machine-executed steps of:

constructing a trainable semantic vector for each data point from the dataset in a multi-dimensional semantic space; [[and]]

applying a clustering process to the constructed trainable semantic vectors to identify similarities between groups of data points within the ~~dataset.~~ dataset; and

providing access to a result of the clustering process;

wherein the trainable semantic vector for each data point from the dataset is constructed by the machine-executed steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories; and

constructing a semantic vector for each data point, wherein each semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories.

30. (Original) The method of Claim 29, wherein the data points correspond to documents.

31. (Original) The method of Claim 29, wherein the step of applying a clustering process comprises the steps:

randomly distributing the data points among a predetermined number of clusters;

determining a cluster center for each cluster;

re-distributing the data points based on the determined cluster centers;

measuring an amount of change in each cluster; and

repeating the steps of determining, re-distributing, and measuring until a predetermined convergence factor has been reached.

32. (Original) The method of Claim 31, wherein:

the step of randomly distributing comprises a step of randomly assigning a fuzzy membership function to each data point; and

the step of re-distributing comprises the step of recalculating the fuzzy membership function for each data point.

33. (Original) The method of Claim 32, further comprising the step of making final cluster assignments based on the fuzzy membership functions.

34. (Original) The method of Claim 33, wherein each data point is assigned to zero or more clusters.

35. (Original) The method of Claim 31, wherein the step of randomly distributing comprises a step of randomly distributing an equal number of data points to each of the predetermined number of clusters.

36. (Original) The method of Claim 31, wherein the predetermined convergence factor is equal to about 0.0001.

37. (Original) The method of Claim 31, wherein the predetermined number of clusters is automatically determined based on the size of the dataset.

38. (Original) The method of Claim 31, wherein the predetermined number of clusters is input by a user.

39. (Original) The method of Claim 31, wherein the step of determining a cluster center comprises a step of constructing an average trainable semantic vector representative of an average value of all datasets within the cluster across all dimensions of the semantic space.

40. (Original) The method of Claim 39, wherein the step of re-distributing comprises a step of assigning the data points to clusters based on the distance from a data point to the nearest cluster center.

Claims 41-61 (Cancelled)

62. (Currently amended) A system for clustering data points from a dataset comprising:

a computer configured to:

construct a trainable semantic vector for each data point from the dataset in a multi-dimensional semantic space;

[[and]]

apply a clustering process to the constructed trainable semantic vectors to identify similarities between groups of data points within the ~~dataset.~~dataset; and

provide access to a result of the clustering process;

wherein the trainable semantic vector for each data point from the dataset is constructed by the machine-executed steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories; and

constructing a semantic vector for each data point, wherein each semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories.

Claims 63-67 (Cancelled)

68. (Currently amended) A computer-readable medium carrying one or more sequences of instructions for clustering data points from a dataset, wherein execution of the one or more sequences of instructions by one or more processors causes the one or more processors to perform the machine-executed steps of:

constructing a trainable semantic vector for each data point from the dataset in a multi-dimensional semantic space; [[and]]

applying a clustering process to the constructed trainable semantic vectors to identify similarities between groups of data points within the ~~dataset.~~dataset; and

providing a result of the clustering process;

wherein the trainable semantic vector for each data point from the dataset is constructed by the machine-executed steps of:

for each data point, identifying a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space;

determining the significance of each data point with respect to the predetermined categories; and

constructing a semantic vector for each data point, wherein each semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories.

Claims 69-74 (Cancelled)